Open Science

Nicolas P. Rougier

OPEN SCIENCE

Nicolas P. Rougier, INRIA
AdaWeek, Paris, 2016

The opposite of open isn't closed
The opposite of open is broken
John Wilbanks, 2007

~~Science~~

Broken Science

PAYWALL

NON-SHARED DATA

CLOSED TOOLS & SOFTWARES

NON-REPRODUCIBLE

ANONYMOUS PEER-REVIEW

PUBLISH OR PERISH

~~Open Science~~

Science

OPEN ACCESS

OPEN DATA

OPEN SOURCE

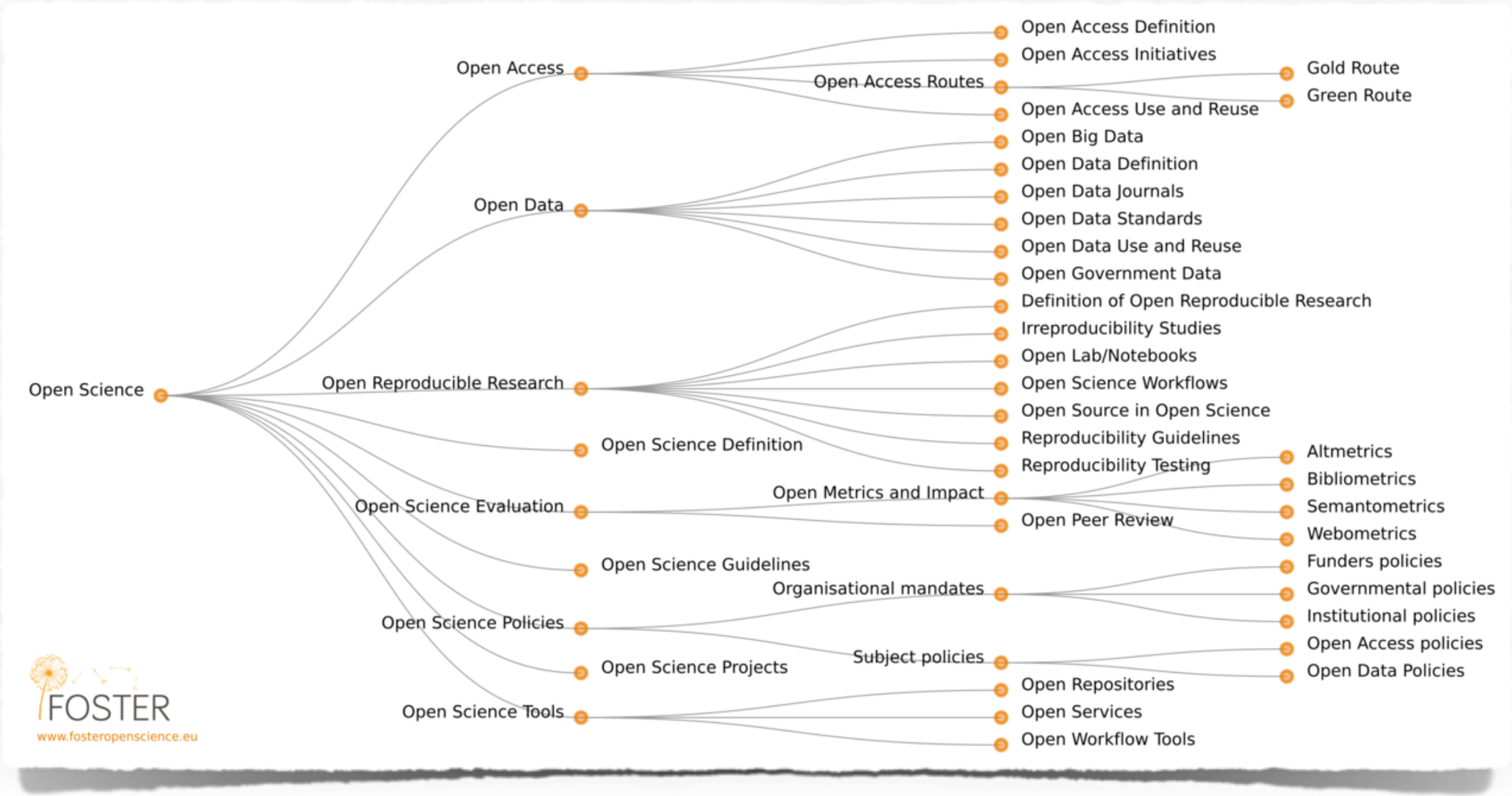
OPEN METHODOLOGY

OPEN PEER-REVIEW

MUCH MORE FUN…

OPEN SCIENCE

Open science is the movement to make scientific research, data and dissemination accessible to all levels of an inquiring society, amateur or professional.



OPEN ACCESS

32¢

Nature 171 (1953)

Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid

J.D. WATSON & F. H. C. CRICK.

38¢

Annalen der Physik, vol. 18, no 13, (1905)

Does the Inertia of a Body Depend upon its Energy-Content

A. EINSTEIN.

Journal of the Proceedings of the Linnean Society of London. Zoology, vol. 3 (1858)

On the Tendency of Species to form Varieties; and on the Perpetuation of Varieties and Species by Natural Means of Selection.

C. DARWIN & A. WALLACE.

38¢

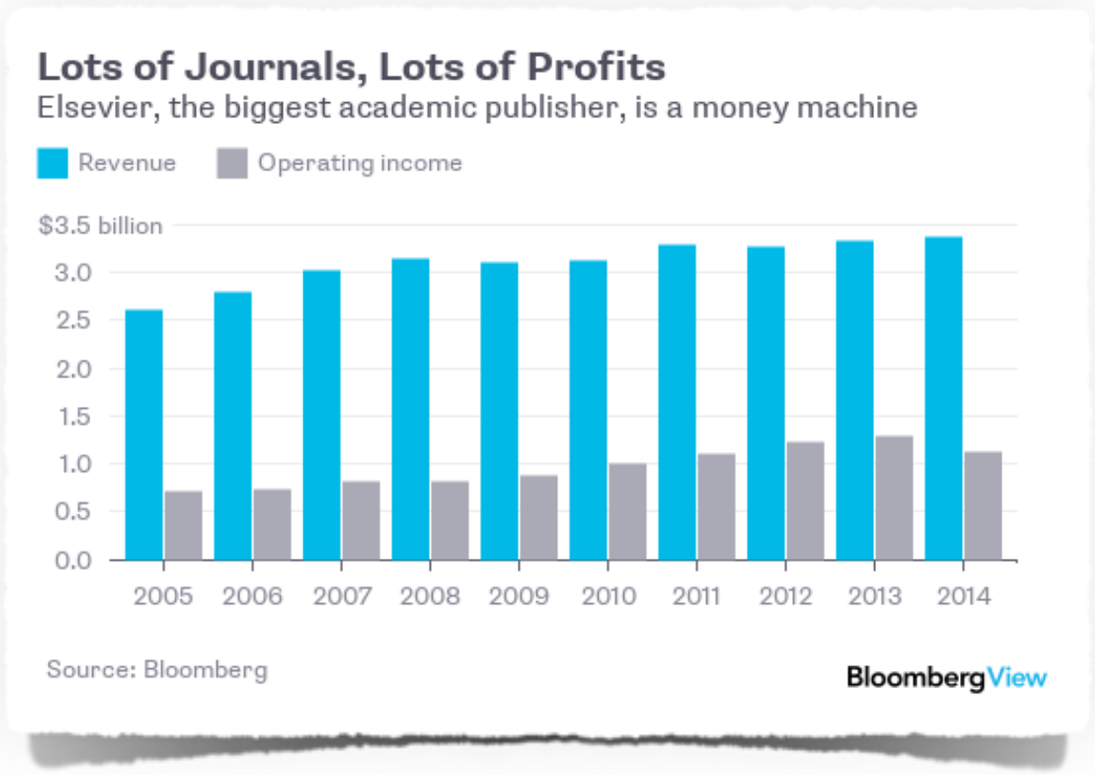
40¢

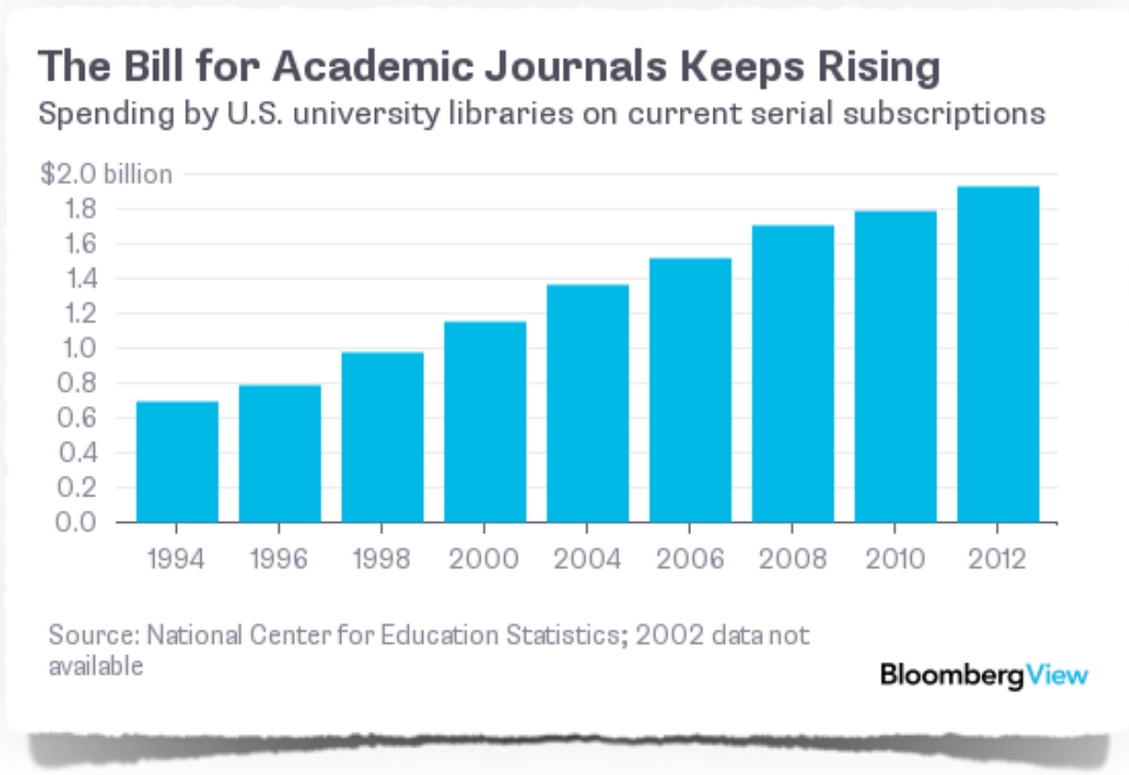
Communications in Mathematical Physics 43 (1975)

Particle creation by black holes

S.W. HAWKING.

OPEN ACCESS





Green access / Gold access / Hybrid access

Nobody pays
(sort of…)

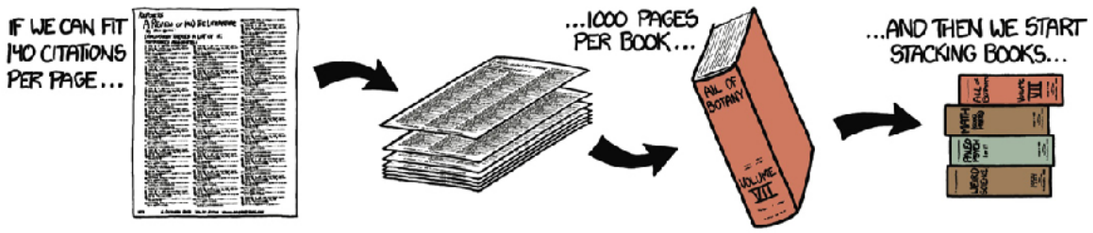
Author pays
(a lot usually)

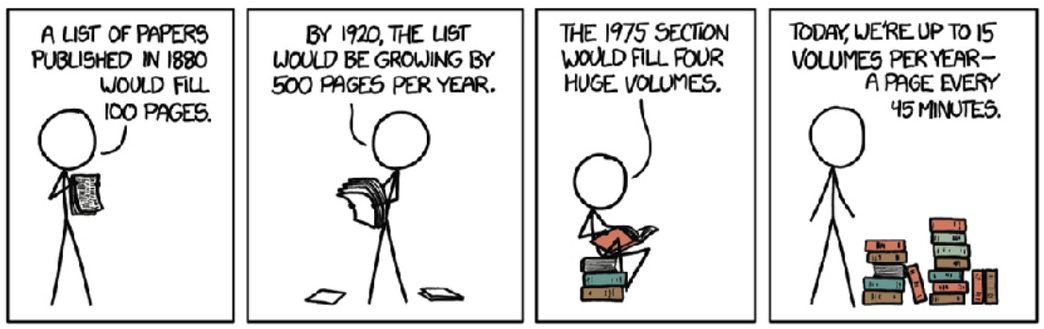
Everybody pays
(what the f…?)

OPEN ACCESS

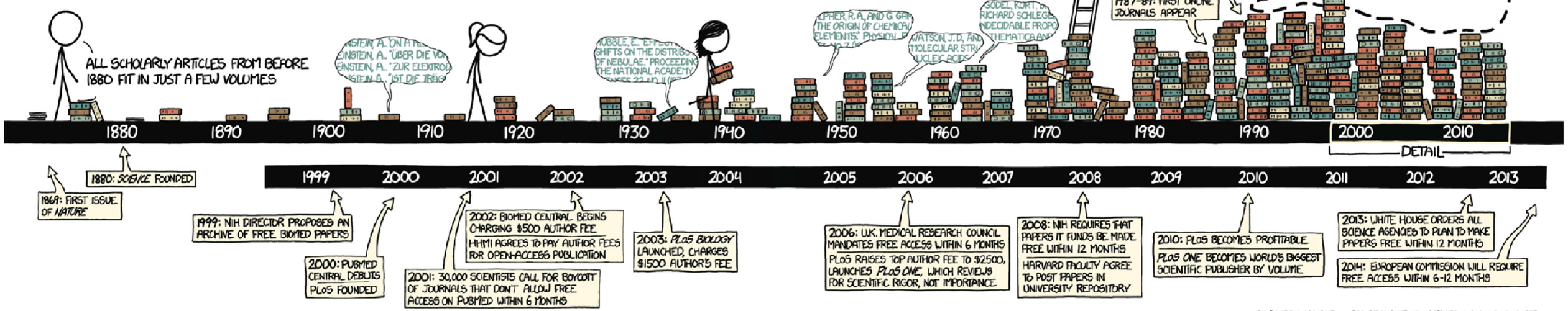
HOW MUCH SCIENCE IS THERE?

SCIENTIFIC PUBLISHING HAS BEEN ACCELERATING—A NEW PAPER IS NOW PUBLISHED ROUGHLY EVERY 20 SECONDS.
LET'S IMAGINE A BIBLIOGRAPHY LISTING *EVERY* SCHOLARLY PAPER EVER WRITTEN.
HOW LONG WOULD IT BE?



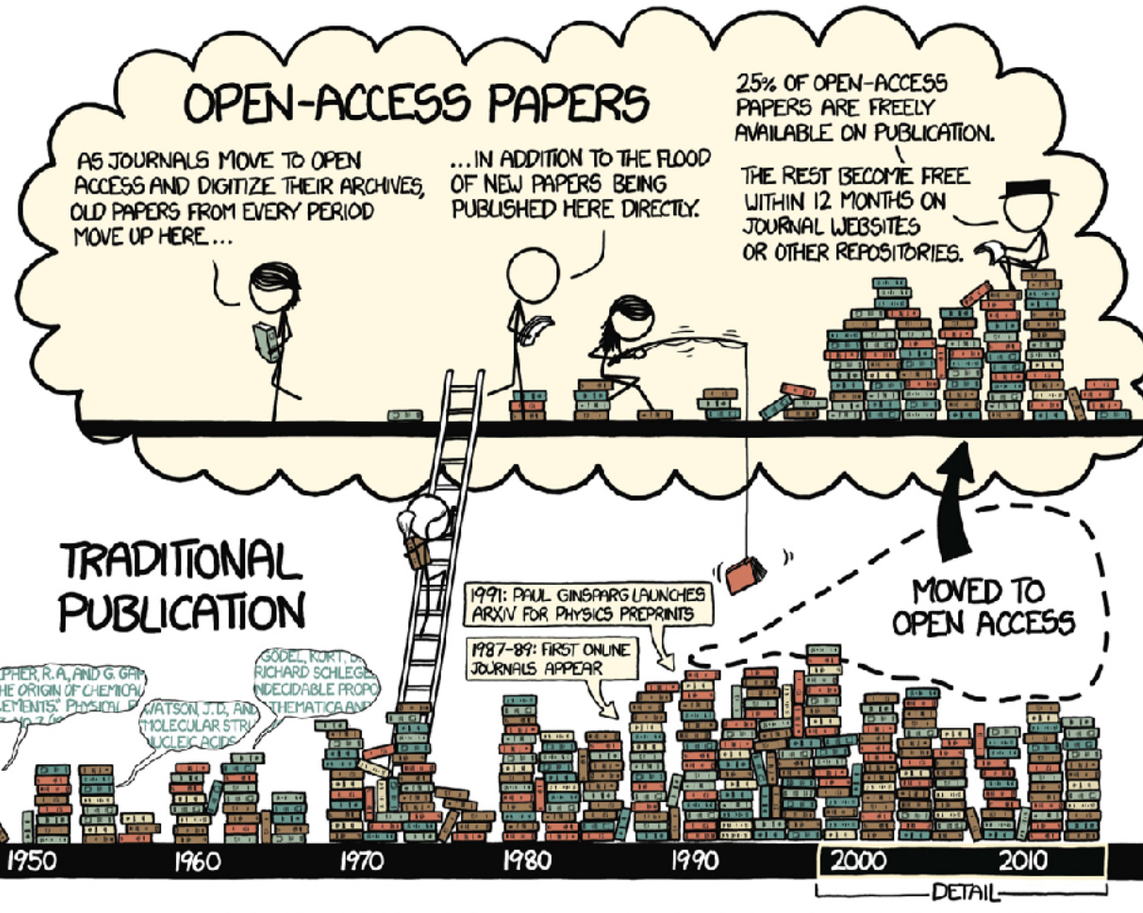


...THIS IS WHAT THE FULL LIST WOULD LOOK LIKE:



HOW OPEN IS IT?

SINCE THE ADVENT OF THE WEB, MUCH OF SCIENTIFIC PUBLISHING HAS BEEN MOVING TO *OPEN ACCESS*.
ACCORDING TO SCIENCE-METRIX, OPEN ACCESS REACHED A "TIPPING POINT" AROUND 2011:
MORE THAN 50% OF NEW RESEARCH IS NOW MADE AVAILABLE FREE ONLINE.



BY RANDALL MUNROE • REPORTING BY JOCELYN KAISER AND DAVID MALAKOFF

OPEN DATA

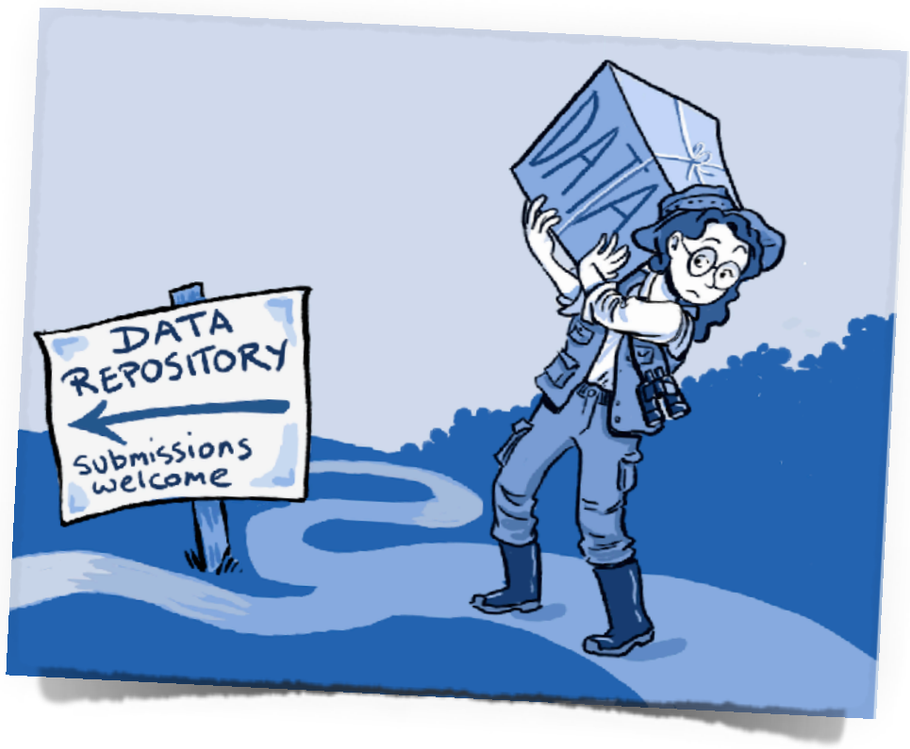
Open data may include non-textual material such as maps, genomes, connectomes, chemical compounds, mathematical and scientific formulae, medical data and practice, bioscience and biodiversity (wikipedia)

“Datasets are available upon request by contacting the corresponding author”
→ How long will the email address last ?

“Code can be found in the appendix”
→ Maybe, maybe not…

“All data are available from my webpage”
→ 404 page not found

“Pseudo-code is given in table 1”
→ And what I’m supposed to do with it ?



OPEN DATA


PUBLISHING BETTER SCIENCE THROUGH BETTER DATA
2016
26TH OCTOBER
@THE WELLCOME COLLECTION
#scidata16
SPRINGER NATURE & WELLCOME
A WARM "WELLCOME" FROM IAIN HRYNASZKIEWICZ
IT'S ALSO OPEN ACCESS WEEK
TODAY IS ALL ABOUT BEING Open
LET'S HAVE THOSE DIFFICULT CONVERSATIONS ABOUT TRANSPARENCY...
A PRACTICAL DEMONSTRATION → sli.do AUDIENCE INTERACTION...
FOR QUESTIONS & (HOPEFULLY) ANSWERS
FLORIAN MARKOWETZ KEYNOTE #1
SELFISH REASONS TO WORK REPRODUCIBLY...
IT HELPS AVOID DISASTER!!
WHO CARES?!
FORGET ABOUT THE HONOURABLE REASONS...
THERE SHOULD BE NO HYPOTHESES IN SCIENCE
YOU SHOULD CARE ABOUT REPRODUCIBILITY ALL THE TIME
IT HELPS PEOPLE SEE IT YOUR WAY
IT HELPS YOU write papers...
IT HELPS TO BUILD YOUR REPUTATION
IT'S GOOD FOR CONTINUITY...
OOOH!
MIND YOUR OWN BUSINESS! ETC
I'VE HEARD ALL THE EXCUSES
BUT REPRODUCIBILITY MEANS SCRUTINY
LONG CVs IS WHAT SCIENCE IS ALL ABOUT...
CV
THESE ARE THE SOFT SKILLS!
BE MORE ORGANISED
LEARN THE TOOLS!
JENNY MOLLY KEYNOTE #2
DATA AND OPEN SCIENCE AS AN EARLY CAREER RESEARCHER
WHAT IS "MY DATA"
EXPLAINING TO PEOPLE WHAT THEY HAVE IS THE FIRST STEP
EVERYONE IS IN THE SAME BOAT...
OPEN DATA BINGO - THERE ARE ALWAYS REASONS NOT TO SHARE...
FIVE STEPS FOR SHARING
1 GET MOTIVATED START EARLY!
2 SHARE LIKE YOU WANT TO...
3 STAY ON TOP OF YOUR DATA
4 MAKE THE MOST OF IT
5 PLAY FAIR
SCRIBING BY @LUDICCREATIVES
LIGHTNING TALKS... 1
ANTICA CULINA
MAKE SENSE OF HUGE AMOUNT OF INFO BY SUMMARISING
HUMANS MAKE MISTAKES WE'RE GOOD AT FOOLING OURSELVES...
FAIR DATA, META-ANALYSIS AND EVOLUTIONARY BIOLOGY. HOW GOOD ARE THESE INGREDIENTS?
NATHAN GOLIGHTLY
HOW WELL CAN WE PREDICT BIOMEDICAL OUTCOMES WITH GENE EXPRESSION PROFILES?
OUR DATA SETS ARE AVAILABLE ON THE OPEN SCIENCE FRAMEWORK
MOIRA NICOLSON
DATA IS OFTEN AS ACCURATE AS YOUR FUEL BILL...
USING ADMIN DATA ON GOVT PLUGIN ELECTRIC VEHICLE GRANTS TO HELP PROMOTE ENERGY SECURITY
YES, WE THREW AWAY A LOT OF DATA SETS!
GUY A. ROULEAU
WE EMAILED PEOPLE ABOUT SWITCHING TARIFF AS MANY NEVER DO!
STOP USERS ALL CHARGING ONE AT SAME TIME
WE HAD TO THINK ABOUT HOW BAD A JOB WE WERE DOING...
WE EMIGRATED OPEN SCIENCE TO GET US QUICKER TO OUR GOAL
MONTREAL NEURO-LOGICAL INSTITUTE
INTELLECTUAL PROPERTY WILL NOT BE FILED
IT WAS A PROCESS OF EDUCATION AND EXCHANGE
MUST BE ETHICALLY ACCEPTABLE PROJECT THE PATIENTS
WHAT CAN WE DO BETTER? IT'S A SOCIAL EXPERIMENT
IT'S THE RIGHT THING TO DO!
THINK OPEN SCIENCE INITIATIVE
LIGHTNING TALKS... 2
ROBERT HONDEN
WE NEED A REPRODUCABLE, TRANSFERRABLE, OPEN PLATFORM FOR 3D VISUALISATION
ANY COMMUNITY CAN USE THIS PLATFORM
3D VISUALISATION OF NANO MATERIALS
WE'VE ESTABLISHED A CHAIN-CANADA-CURA NOW!
WE'VE GOT THE US SURROUNDED!
RACHEL HARDING
OPEN LAB NOTEBOOKS PROS & CONS
LOTS OF PEOPLE SEEING THE DATA
I'M SHARING ALL OF MY DATA ON HUNTINGDON'S DISEASE!
UTILISING SOCIAL MEDIA
IT'S REAL TIME SO THERE WILL BE SOME SPELLING MISTAKES!!
ALAN EVANS
CRAIN & THE CANADIAN BRAIN INITIATIVE
DATA CAN BE ANYWHERE
THE COMPUTE CAN BE ANYWHERE
USERS CAN BE ANYWHERE
JO BARRATT
CONTINUOUS DATA VALIDATION IN SCIENTIFIC RESEARCH
MANAGING DATA CAN BE EXPENSIVE
WE NEED TO REMOVE THE FRICTION IN WORKING WITH DATA
PACKAGE DATA IS BETTER
GIRIJA GOYAL & JAMES AKIN
WE WANT TO GIVE DATA CONTEXT
THIS MAKES IT SEAMLESS
CHALLENGE IS TO GET SCIENTISTS TO ADOPT IT
REFIGURE: COMPARING, ANNOTATING & GIVING CREDIT FOR REPRODUCED DATA
TRACEBACK... INFORMATION MUST BE TRACEABLE
IT'S A RISK AVERSE CULTURE IF AN ORG DOES NOT SHARE NO ONE GETS IN TROUBLE
ALL DATA MUST BE SUBJECT TO THE SAME SCRUTINY
MAKE YOUR DATA DISCOVERABLE
CAN'T KEEP ALL DATA... SOME MUST BE THROWN AWAY
BUT BE INTELLIGENT ABOUT IT
KNOW... WHAT YOU HAVE GOT RID OF & WHAT CAN BE RE-CREATED (OR NOT)
...AND FINALLY
PANEL DISCUSSION
BENEFITS AND UNINTENDED CONSEQUENCES OF TRANSPARENCY
EMMA GANLEY
JON PISTER
STEPHAN LEWAN-DOWSKY
ALISON MITCHELL
PERMUTATING....
WE NOW ASK WHERE CAN THE DATA BE FOUND?
WE DON'T ASK FOR ALL DATA
THE CULTURE IS ON THE CUSP OF CHANGE
LET'S PUSH FOR MORE CHANGE
BE WARY OF WHO WANTS THE DATA (EG TOBACCO INDUSTRY)
THINK ABOUT THE DATA JOURNEY
RESEARCH UNITS LIKE SCHOOLS SHOULD HAVE AN OPEN DATA REPRESENTATIVE
FIND YOUR GURUS AND FOLLOW THEM CLOSELY
LET'S NOT IMPOSE REQUIREMENTS ON PEOPLE
THERE HAS TO BE ORGANISATIONAL STRUCTURE, POLICY ETC
"NAKED" GOVERNMENT (ACCESS TO EVERYTHING) IS UNDEMOCRATIC
KEVIN ASHLEY KEYNOTE #3
RESEARCH DATA MANAGEMENT
HOW WILL OTHERS USE MY DATA WHEN I'VE DONE WITH IT?
WE'VE ALWAYS USED THE WISDOM OF THE ANCIENTS ONE DAY WE'LL BE THE ANCIENTS (LET'S MAKE DATA STILL USEFUL!)
FIND A HOME FOR YOUR DATA
EQUIP YOURSELF WITH THE RIGHT SKILLS
USE A DATA MANAGEMENT PLAN...
SHORT SIMPLE SPECIFIC
OTHERS CAN BUILD ON THIS
THINK OUTSIDE THE BOUNDARIES
SITUATION IS NOT PERFECT BUT HEADING IN THE RIGHT DIRECTION
ANDREW HUFTON
KEY NOTE #4
THE IMPORTANCE OF RESEARCH DATA FOR GETTING PUBLISHED
YOU HAVE TO SHOW YOUR EVIDENCE IF YOU ARE PUBLISHING
WE TEND TO TREAT DATA SHARING LIKE BROCCOLI... WE KNOW IT'S GOOD FOR US BUT NO ONE LIKES IT
THERE'S A LOT MORE BIG DATA NOW
WE HAVE CLEAR PREFERENCE FOR LARGE IMAGES
MAJORITY OF RESEARCHERS ARE SHARING THEIR DATA
FAIR DATA
FINDABLE
ACCESSIBLE
INTEROPERABLE
RE-USABLE
"SCIENTIFIC DATA" - WE'RE TOTALLY OPEN ACCESS
PRESERVE YOUR DATA
PUT YOUR DATA IN A...
GOOD REPOSITORY
LOOK AT GUIDELINES, AT WHAT YOU OUGHT TO BE DOING - NOTHING NEW HERE
WE NEED TO DEAL WITH THIS "DATA FACT" NEED
ENCOURAGE PEOPLE TO USE IT...

OPEN DATA

Open data is the idea that some data should be freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control.

FAIR Data : Findable, Accessible, Interoperable, Re-Usable

A second concern held by some is that a new class of research person will emerge — people who had nothing to do with the design and execution of the study but use another group's data for their own ends, possibly stealing from the research productivity planned by the data gatherers, or even **use the data to try to disprove** what the original investigators had posited.There is concern among some front-line researchers that the system will be taken over by what some researchers have characterized as **research parasites**

The New England Journal of Medicine (2016)

OPEN METHODOLOGY

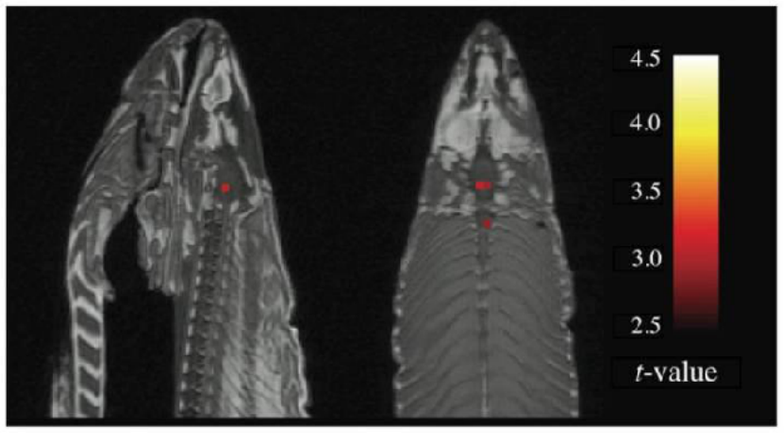
Debt

FAQ: Reinhart, Rogoff, and the Excel Error That Changed History

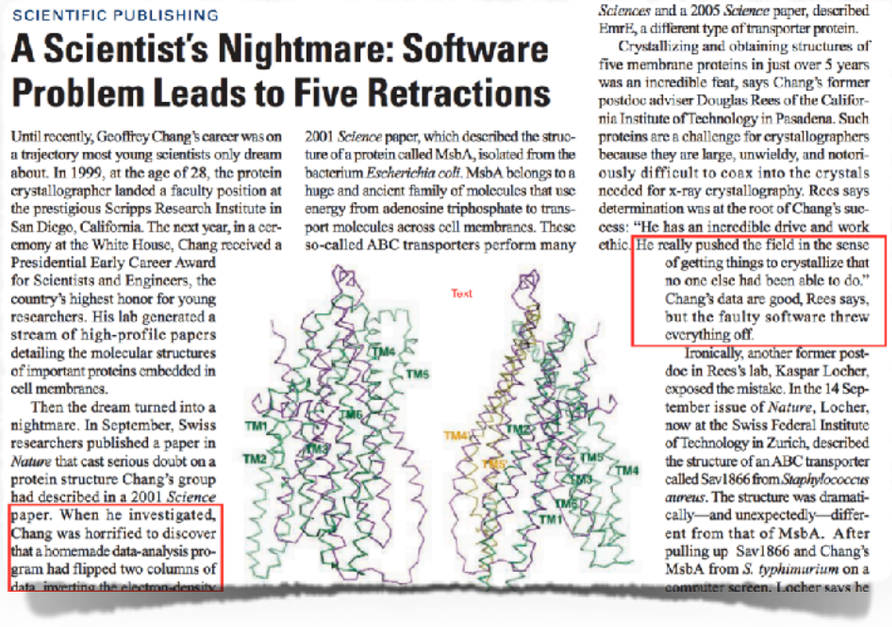
By Peter Coy | April 18, 2013

Photograph by Gregor Schuster

Harvard University economists Carmen Reinhart and Kenneth Rogoff have acknowledged making a spreadsheet calculation mistake in a 2010 research paper, "Growth in a Time of Debt" (PDF), which has been widely cited to justify budget-



A *t*-contrast was used to test for regions with significant BOLD signal change during the photo condition compared to rest. The parameters for this comparison were $t(131) > 3.15$, p(uncorrected) < 0.001, 3 voxel extent threshold.



SCIENTIFIC PUBLISHING

A Scientist's Nightmare: Software Problem Leads to Five Retractions

Until recently, Geoffrey Chang's career was on a trajectory most young scientists only dream about. In 1999, at the age of 28, the protein crystallographer landed a faculty position at the prestigious Scripps Research Institute in San Diego, California. The next year, in a ceremony at the White House, Chang received a Presidential Early Career Award for Scientists and Engineers, the country's highest honor for young researchers. His lab generated a stream of high-profile papers detailing the molecular structures of important proteins embedded in cell membranes.

Then the dream turned into a nightmare. In September, Swiss researchers published a paper in *Nature* that cast serious doubt on a protein structure Chang's group had described in a 2001 *Science* paper. When he investigated, Chang was horrified to discover that a homemade data-analysis program had flipped two columns of data, inverting the electron-density

2001 *Science* paper, which described the structure of a protein called MsbA, isolated from the bacterium *Escherichia coli*. MsbA belongs to a huge and ancient family of molecules that use energy from adenosine triphosphate to transport molecules across cell membranes. These so-called ABC transporters perform many

Sciences and a 2005 *Science* paper, described EmrE, a different type of transporter protein.

Crystallizing and obtaining structures of five membrane proteins in just over 5 years was an incredible feat, says Chang's former postdoc adviser Douglas Rees of the California Institute of Technology in Pasadena. Such proteins are a challenge for crystallographers because they are large, unwieldy, and notoriously difficult to coax into the crystals needed for x-ray crystallography. Rees says determination was at the root of Chang's success: "He has an incredible drive and work ethic. He really pushed the field in the sense of getting things to crystallize that no one else had been able to do." Chang's data are good, Rees says, but the faulty software threw everything off.

Ironically, another former postdoc in Rees's lab, Kaspar Locher, exposed the mistake. In the 14 September issue of *Nature*, Locher, now at the Swiss Federal Institute of Technology in Zurich, described the structure of an ABC transporter called Sav1866 from *Staphylococcus aureus*. The structure was dramatically—and unexpectedly—different from that of MsbA. After pulling up Sav1866 and Chang's MsbA from *S. typhimurium* on a computer screen, Locher says he

OPEN METHODOLOGY

I have heard from graduate students opting out of academia, assistant professors afraid to come up for tenure, mid-career people wondering how to protect their labs, and senior faculty retiring early, all because of **methodological terrorism**.

APS Observer (2016)

The (ugly) hacking way:

```
p=0.052, click, click, click, p=0.049, save
```

The (bad) excel way:

```
click, click, select, click, click, move, del, click, re-
click, save, unclick, ctrl-z, ctrl-z, click, enter, click,
click, move, click, re-click, save
```

The (good) reproducible way:

```
>>> data = load(“data.txt”)
>>> mean, std = data.mean(), data.std()
>>> file = file.open(“analysis.txt”)
>>> write(“mean: %f, std: %f” % (mean, std))
>>> exit()
$ git commit -a -m “Computed results from 24/11/2016”
```

OPEN METHODOLOGY

PRL 116, 061102 (2016) Selected for a Viewpoint in *Physics* PHYSICAL REVIEW LETTERS week ending 12 FEBRUARY 2016

Observation of Gravitational Waves from a Binary Black Hole Merger

B. P. Abbott *et al.**
(LIGO Scientific Collaboration and Virgo Collaboration)
(Received 21 January 2016; published 11 February 2016)

On September 14, 2015 at 09:50:45 UTC the two detectors of the Laser Interferometer Gravitational-Wave Observatory simultaneously observed a transient gravitational-wave signal. The signal sweeps upwards in frequency from 35 to 250 Hz with a peak gravitational-wave strain of 1.0×10^{-21}. It matches the waveform predicted by general relativity for the inspiral and merger of a pair of black holes and the ringdown of the resulting single black hole. The signal was observed with a matched-filter signal-to-noise ratio of 24 and a false alarm rate estimated to be less than 1 event per 203 000 years, equivalent to a significance greater than 5.1σ. The source lies at a luminosity distance of 410^{+160}_{-180} Mpc corresponding to a redshift $z = 0.09^{+0.03}_{-0.04}$. In the source frame, the initial black hole masses are $36^{+5}_{-4} M_\odot$ and $29^{+4}_{-4} M_\odot$, and the final black hole mass is $62^{+4}_{-4} M_\odot$, with $3.0^{+0.5}_{-0.5} M_\odot c^2$ radiated in gravitational waves. All uncertainties define 90% credible intervals. These observations demonstrate the existence of binary stellar-mass black hole systems. This is the first direct detection of gravitational waves and the first observation of a binary black hole merger.

DOI: 10.1103/PhysRevLett.116.061102

I. INTRODUCTION

In 1916, the year after the final formulation of the field equations of general relativity, Albert Einstein predicted the existence of gravitational waves. He found that the linearized weak-field equations had wave solutions: transverse waves of spatial strain that travel at the speed of light, generated by time variations of the mass quadrupole moment of the source [1,2]. Einstein understood that gravitational-wave amplitudes would be remarkably small; moreover, until the Chapel Hill conference in 1957 there was significant debate about the physical reality of gravitational waves [3].

Also in 1916, Schwarzschild published a solution for the field equations [4] that was later understood to describe a black hole [5,6], and in 1963 Kerr generalized the solution to rotating black holes [7]. Starting in the 1970s theoretical work led to the understanding of black hole quasinormal modes [8–10], and in the 1990s higher-order post-Newtonian calculations [11] preceded extensive analytical studies of relativistic two-body dynamics [12,13]. These advances, together with numerical relativity breakthroughs in the past decade [14–16], have enabled modeling of binary black hole mergers and accurate predictions of their gravitational waveforms. While numerous black hole candidates have now been identified through electromagnetic observations [17–19], black hole mergers have not previously been observed.

The discovery of the binary pulsar system PSR B1913+16 by Hulse and Taylor [20] and subsequent observations of its energy loss by Taylor and Weisberg [21] demonstrated the existence of gravitational waves. This discovery, along with emerging astrophysical understanding [22], led to the recognition that direct observations of the amplitude and phase of gravitational waves would enable studies of additional relativistic systems and provide new tests of general relativity, especially in the dynamic strong-field regime.

Experiments to detect gravitational waves began with Weber and his resonant mass detectors in the 1960s [23], followed by an international network of cryogenic resonant detectors [24]. Interferometric detectors were first suggested in the early 1960s [25] and the 1970s [26]. A study of the noise and performance of such detectors [27], and further concepts to improve them [28], led to proposals for long-baseline broadband laser interferometers with the potential for significantly increased sensitivity [29–32]. By the early 2000s, a set of initial detectors was completed, including TAMA 300 in Japan, GEO 600 in Germany, the Laser Interferometer Gravitational-Wave Observatory (LIGO) in the United States, and Virgo in Italy. Combinations of these detectors made joint observations from 2002 through 2011, setting upper limits on a variety of gravitational-wave sources while evolving into a global network. In 2015, Advanced LIGO became the first of a significantly more sensitive network of advanced detectors to begin observations [33–36].

A century after the fundamental predictions of Einstein and Schwarzschild, we report the first direct detection of gravitational waves and the first direct observation of a binary black hole system merging to form a single black hole. Our observations provide unique access to the

*Full author list given at the end of the article.

SIGNAL PROCESSING WITH GW150914 OPEN DATA

Welcome! This ipython notebook (or associated python script GW150914_tutorial.py) will go through some typical signal processing tasks on strain time-series data associated with the LIGO GW150914 data release from the LIGO Open Science Center (LOSC):

- https://losc.ligo.org/events/GW150914/
- View the tutorial as a web page - https://losc.ligo.org/s/events/GW150914/GW150914_tutorial.html
- Download the tutorial as a python script - https://losc.ligo.org/s/events/GW150914/GW150914_tutorial.py
- Download the tutorial as iPython Notebook - https://losc.ligo.org/s/events/GW150914/GW150914_tutorial.ipynb

To begin, download the ipython notebook, readligo.py, and the data files listed below, into a directory / folder, then run it. Or you can run the python script GW150914_tutorial.py. You will need the python packages: numpy, scipy, matplotlib, h5py.

On Windows, or if you prefer, you can use a python development environment such as Anaconda (https://www.continuum.io/why-anaconda) or Enthought Canopy (https://www.enthought.com/products/canopy/).

Questions, comments, suggestions, corrections, etc: email losc@ligo.org

v20160208b

Intro to signal processing

This tutorial assumes that you know python well enough.

If you know how to use "ipython notebook", use the GW150914_tutorial.ipynb file. Else, you can use the GW150914_tutorial.py script.

This tutorial assumes that you know a bit about signal processing of digital time series data (or want to learn!). This includes power spectral densities, spectrograms, digital filtering, whitening, audio manipulation. This is a vast and complex set of topics, but we will cover many of the basics in this tutorial.

If you are a beginner, here are some resources from the web:

- http://101science.com/dsp.htm
- https://georgemdallas.wordpress.com/2014/05/14/wavelets-4-dummies-signal-processing-fourier-transforms-and-heisenberg/
- https://en.wikipedia.org/wiki/Signal_processing
- https://en.wikipedia.org/wiki/Spectral_density
- https://en.wikipedia.org/wiki/Spectrogram
- http://greenteapress.com/thinkdsp/
- https://en.wikipedia.org/wiki/Digital_filter

And, well, lots more - google it!

Download the data

- Download the data files from LOSC:
- We will use the hdf5 files, both H1 and L1, with durations of 32 and 4096 seconds around GW150914, sampled at 16384 and 4096 Hz :
  - https://losc.ligo.org/s/events/GW150914/H-H1_LOSC_4_V1-1126259446-32.hdf5
  - https://losc.ligo.org/s/events/GW150914/L-L1_LOSC_4_V1-1126259446-32.hdf5
  - https://losc.ligo.org/s/events/GW150914/H-H1_LOSC_16_V1-1126259446-32.hdf5
  - https://losc.ligo.org/s/events/GW150914/L-L1_LOSC_16_V1-1126259446-32.hdf5
  - https://losc.ligo.org/s/events/GW150914/GW150914_4_NR_waveform.txt
- Download the python functions to read the data: https://losc.ligo.org/s/sample_code/readligo.py
- From a unix/mac-osx command line, you can use wget; for example,
  - wget https://losc.ligo.org/s/events/GW150914/H-H1_LOSC_4_V1-1126257414-4096.hdf5
- Put these files in your current directory / folder. Don't mix any other LOSC data files in this directory, or readligo.py may get confused.

Here,

- "H-H1" means that the data come from the LIGO Hanford Observatory site and the LIGO "H1" detector;
- the "4" means the strain time-series data are (down-)sampled from 16384 Hz to 4096 Hz;
- the "V1" means version 1 of this data release;
- "1126257414-4096" means the data starts at GPS time 1126257414 (Mon Sep 14 09:16:37 GMT 2015), duration 4096 seconds;
  - NOTE: GPS time is number of seconds since Jan 6, 1980 GMT. See http://www.oc.nps.edu/oc2902w/gps/timsys.html or https://losc.ligo.org/gps/
- the filetype "hdf5" means the data are in hdf5 format: https://www.hdfgroup.org/HDF5/

Note that the the 4096 second long files at 16384 Hz sampling rate are fairly big files (125 MB). You won't need them for this tutorial:

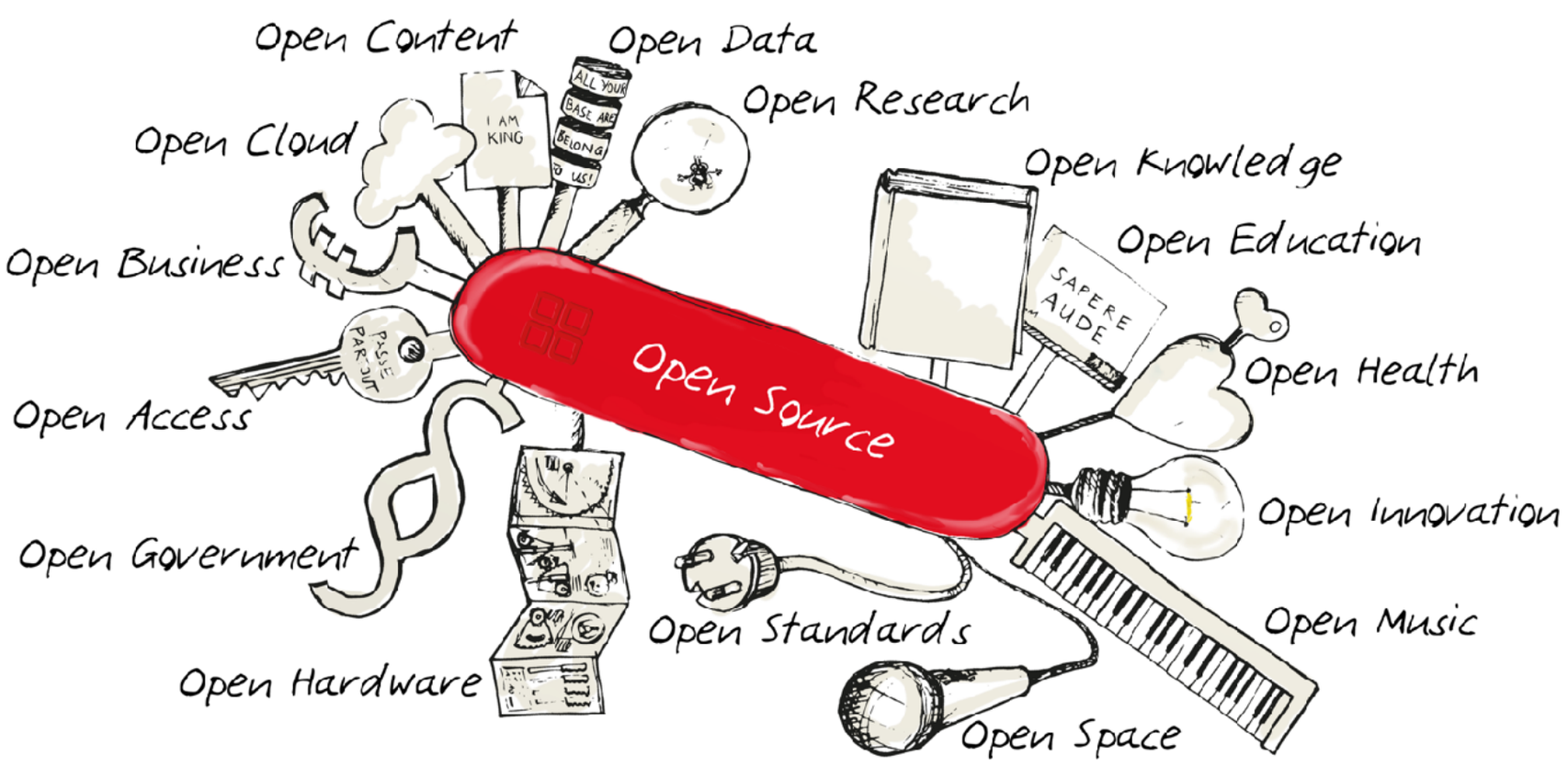
- https://losc.ligo.org/s/events/GW150914/H-H1_LOSC_4_V1-1126257414-4096.hdf5
- https://losc.ligo.org/s/events/GW150914/L-L1_LOSC_4_V1-1126257414-4096.hdf5
- https://losc.ligo.org/s/events/GW150914/H-H1_LOSC_16_V1-1126257414-4096.hdf5

OPEN SOURCE

WHAT ELSE ?

I consider that the Golden Rule requires that if I like a program I must share it with other people who like it. Software sellers want to divide the users and conquer them, making each user agree not to share with others. I refuse to break solidarity with other users in this way.

GNU Manifesto, Richard M. Stallman, 1985

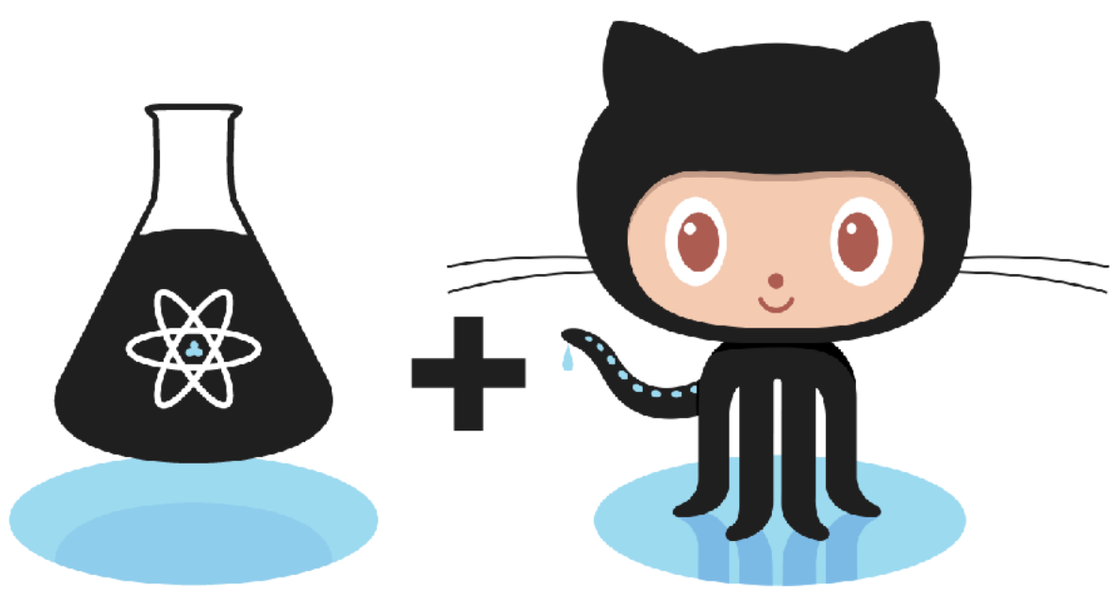


OPEN SOURCE

WHAT ELSE ?

Margaret Hamilton
Director of the Software Engineering Division
https://github.com/chrislgarry/Apollo-11

github : the open-source way to host, create and curate knowledge



Software Heritage

OPEN REVIEW

Nature 333 (1988)

Human basophil degranulation triggered by very dilute antiserum against IgE

E. DAVENAS, F. BEAUVAIS, J. AMARA, M. OBERBAUM, B. ROBINZON, A. MIADONNAI, A. TEDESCHI, B. POMERANZ, P. FORTNER, P. BELON, J. SAINTE-LAUDY, B. POITEVIN & J. BENVENISTE

Social Text, 46/47 (1996)

Transgressing the Boundaries: Toward a Transformative Hermeneutics of Quantum Gravity

A. SOKAL

International Journal of Advanced Computer Technology (2014)

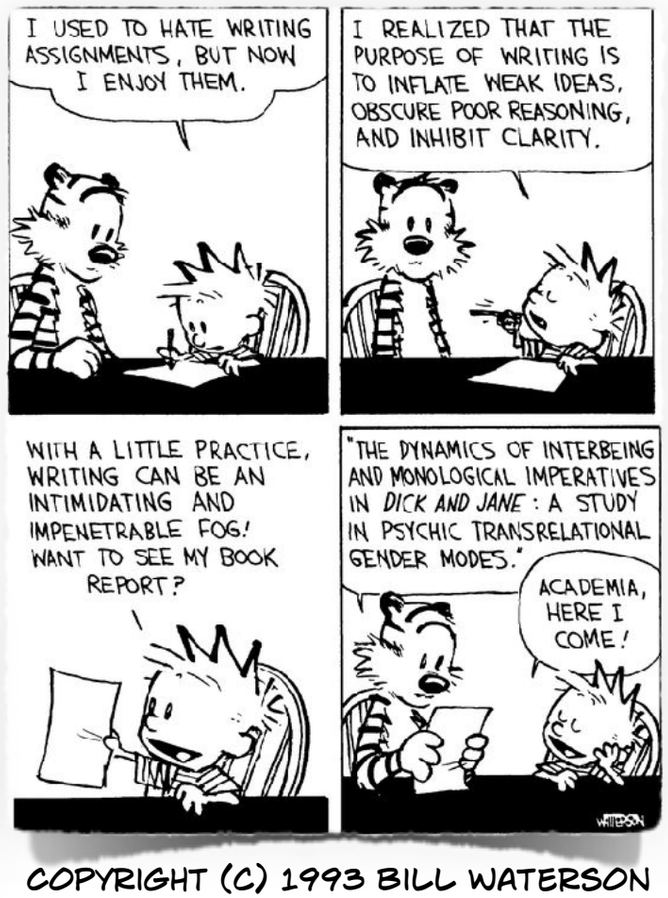
Get Me Off Your Fucking Mailing List

D. MAZIÈRES & E. KOHLER

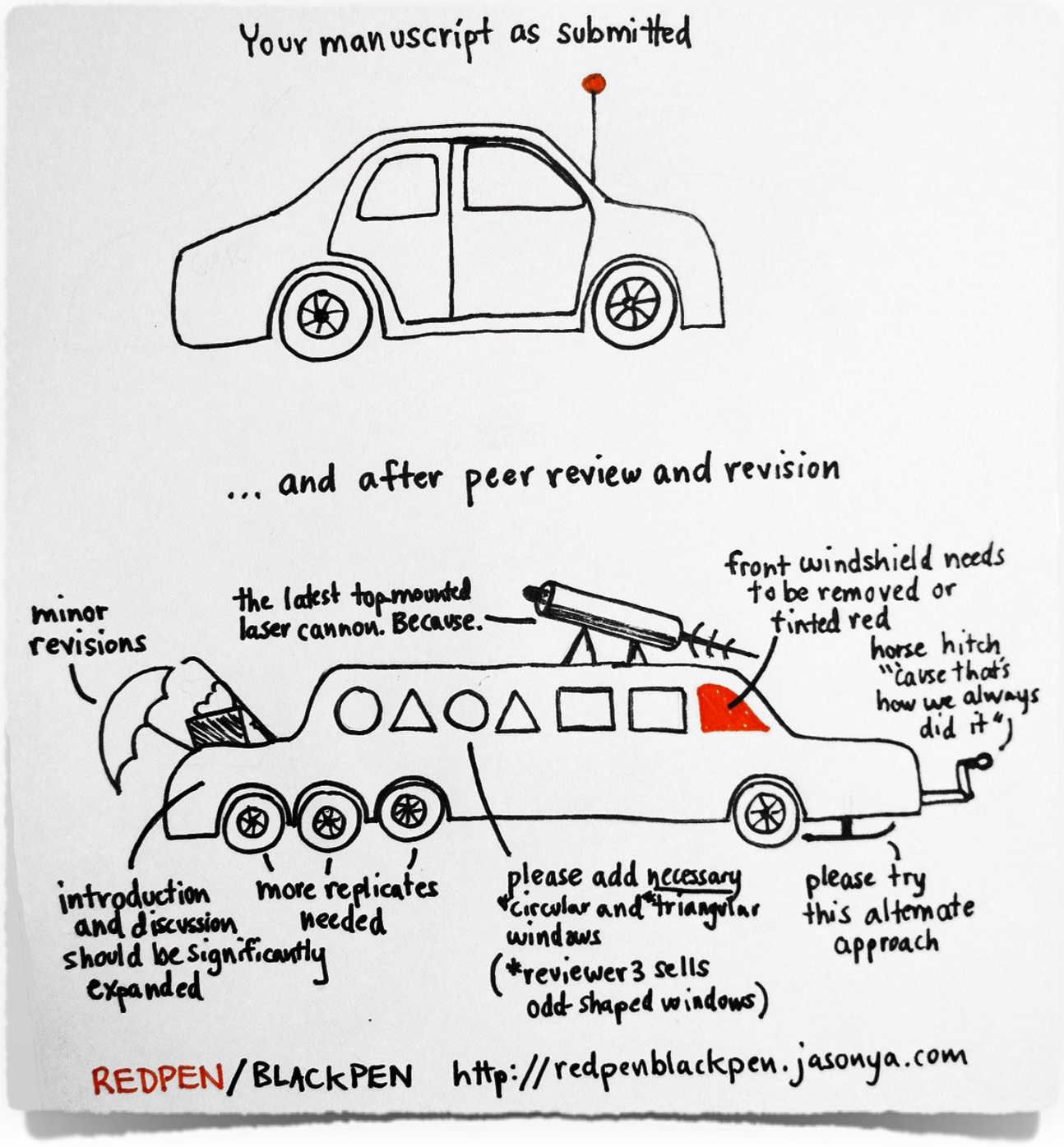
Unpublished (1993)

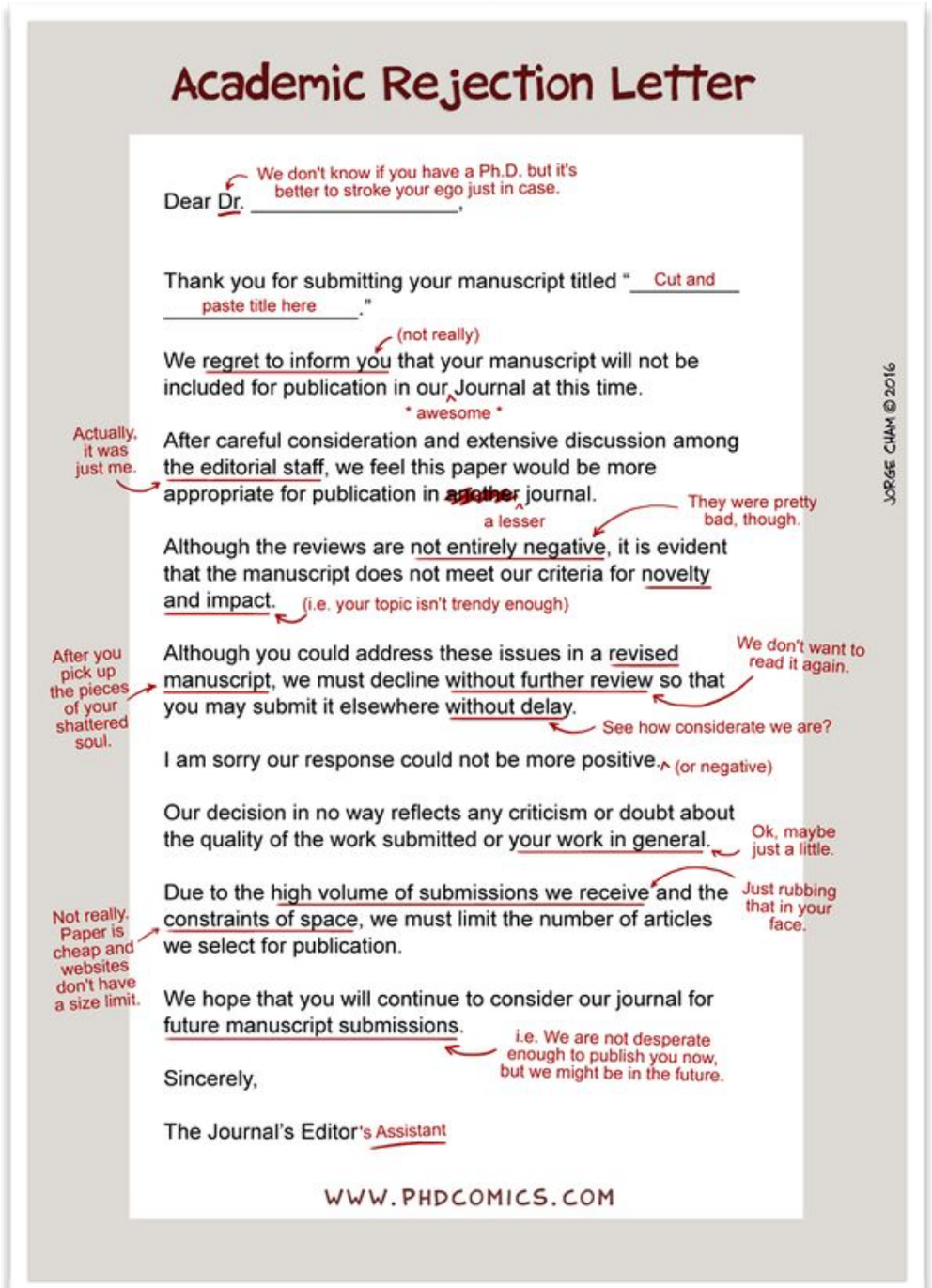
The dynamics of interbeings and monological imperatives in *Dick and Jane*: a study in psychic translational gender modes

CALVIN & HOBBES



OPEN REVIEW


Your manuscript as submitted
... and after peer review and revision
minor revisions
the latest top-mounted laser cannon. Because.
front windshield needs to be removed or tinted red
horse hitch "cause that's how we always did it")
introduction and discussion should be significantly expanded
more replicates needed
please add necessary *circular and *triangular windows (*reviewer 3 sells odd-shaped windows)
please try this alternate approach
REDPEN/BLACKPEN http://redpenblackpen.jasonya.com


Academic Rejection Letter
Dear Dr. _____,
We don't know if you have a Ph.D. but it's better to stroke your ego just in case.
Thank you for submitting your manuscript titled "Cut and paste title here."
We regret to inform you (not really) that your manuscript will not be included for publication in our Journal at this time.
* awesome *
Actually, it was just me.
After careful consideration and extensive discussion among the editorial staff, we feel this paper would be more appropriate for publication in another (a lesser) journal.
They were pretty bad, though.
Although the reviews are not entirely negative, it is evident that the manuscript does not meet our criteria for novelty and impact. (i.e. your topic isn't trendy enough)
After you pick up the pieces of your shattered soul.
Although you could address these issues in a revised manuscript, we must decline without further review so that you may submit it elsewhere without delay.
We don't want to read it again.
See how considerate we are?
I am sorry our response could not be more positive. (or negative)
Our decision in no way reflects any criticism or doubt about the quality of the work submitted or your work in general.
Ok, maybe just a little.
Not really. Paper is cheap and websites don't have a size limit.
Due to the high volume of submissions we receive and the constraints of space, we must limit the number of articles we select for publication.
Just rubbing that in your face.
We hope that you will continue to consider our journal for future manuscript submissions.
i.e. We are not desperate enough to publish you now, but we might be in the future.
Sincerely,
The Journal's Editor's Assistant
WWW.PHDCOMICS.COM
JORGE CHAM © 2016

OPEN REVIEW

ogrisel commented on Jul 23 — ReScience organization member

Ok I trained the "wide 2 4" configuration and could reach 0.957476 final validation accuracy. However the `weights` folder did not exist (probably because git does not checkout empty folders) so saving the weights failed. You might want to make sure to create the folders if they don't exists before writing files in this script :)

```
iter: 198 | TL: 0.192 | VL: 0.202 | Vacc: 0.956 | Ratio: 0.95 | Time: 157.3
iter: 199 | TL: 0.192 | VL: 0.2 | Vacc: 0.955 | Ratio: 0.96 | Time: 157.3
Final Acc: 0.957476
Traceback (most recent call last):
  File "train_nn.py", line 133, in <module>
    f = gzip.open('data/weights/%s%d_resnet.pklz'%(variant,depth), 'wb')
  File "/usr/lib/python2.7/gzip.py", line 34, in open
    return GzipFile(filename, mode, compresslevel)
  File "/usr/lib/python2.7/gzip.py", line 94, in __init__
    fileobj = self.myfileobj = __builtin__.open(filename, mode or 'rb')
IOError: [Errno 2] No such file or directory: 'data/weights/wide2_resnet.pklz'
```

FlorianMuellerklein commented on Jul 23 • edited

@ogrisel Oh no, I'm really sorry about that. I'll add that to my next update.

EDIT: Updated to include python package versions, commands to reproduce. train_nn.py now creates the necessary folders if they don't already exist.

Updated README with software version numbers and training script to c… ⋯ 0134558

Projects

None yet

Labels

01 - Request

02 - Review

Engineering Science

Python

Milestone

No milestone

Assignees

emmanuelle

5 participants

Notifications

Unsubscribe

You're receiving notifications because you commented.

Lock conversation

OPEN EDUCATION

Open Education Week
2017
March 27 - 31

Open education is a collective term to describe institutional practices and programmatic initiatives that broaden access to the learning and training traditionally offered through formal education systems.

NEW PLAYERS

ORCID (Open Researcher and Contributor ID) is a nonproprietary alphanumeric code to uniquely identify scientific and other academic authors and contributors

Figshare is an online digital repository where researchers can preserve and share their research outputs, including figures, datasets, images, and videos.

Zenodo is a research data repository created by OpenAIRE and CERN to provide a place for researchers to deposit datasets.

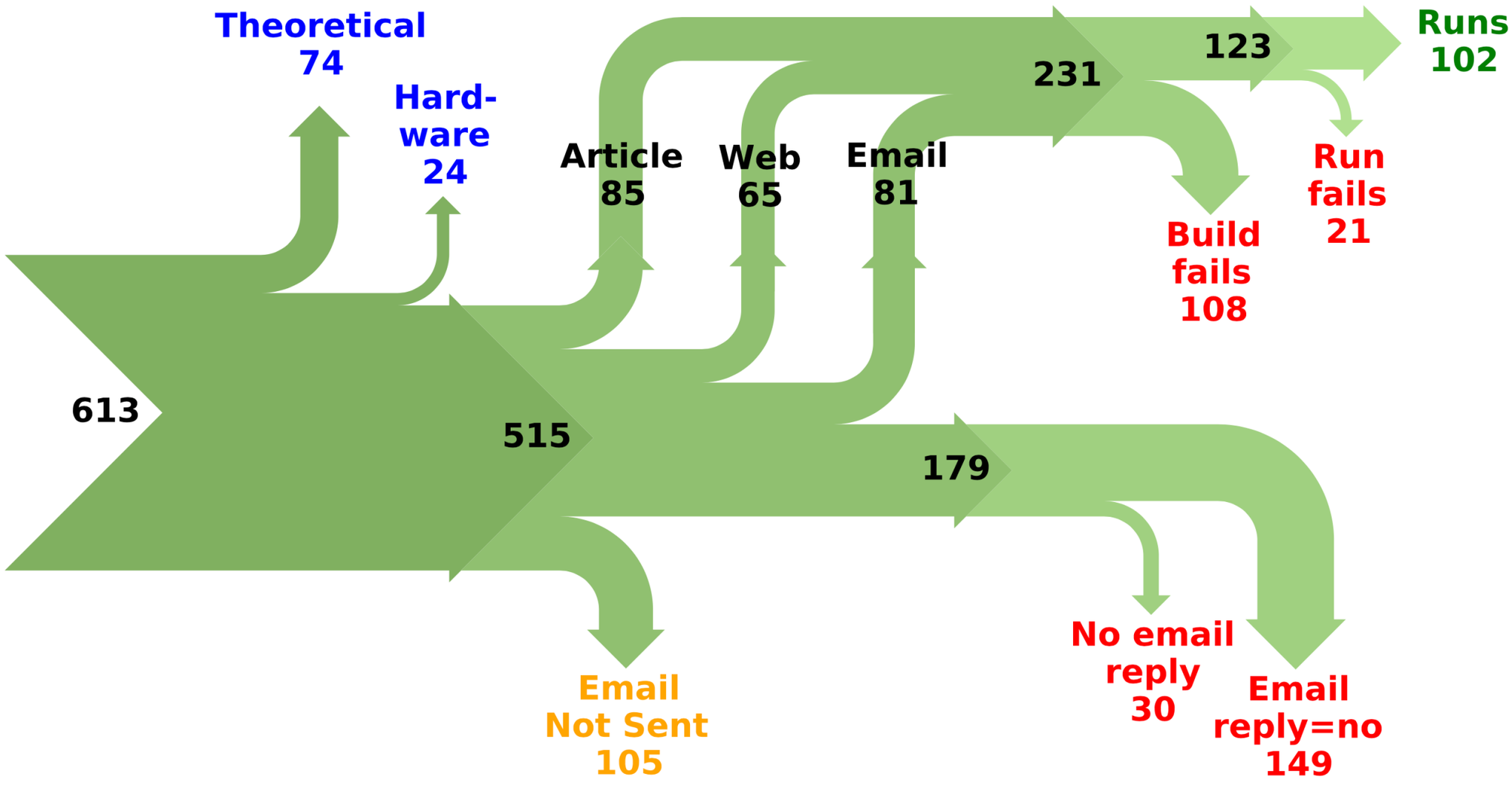
ArXiv is a repository of electronic preprints of scientific papers in the fields of mathematics, physics, astronomy, computer science, quantitative biology, statistics, and quantitative finance, which can be accessed online.

GitHub is a web-based Git repository hosting service that offers all of the distributed version control and source code management functionality of git as well as adding its own features.

And many more to discover (RIO journal, F1000, The Winnover, Jupyter project, Software & Data carpentry, etc.)

SUCCESS STORY

(SELF-ADVERTISEMENT)



Measuring Reproducibility in Computer Systems Research

Colbert et al. (2014)

SUCCESS STORY

(SELF-ADVERTISEMENT)

ReScience

Reproducible science is good. Replicated science is better.

OPEN DATA

OPEN SOURCE

OPEN PEER-REVIEW

OPEN (GREEN) ACCESS

NO "BUZZ" BARRIER

COMMUNITY SUPPORTED

0€ BUDGET

SUCCESS STORY

(SELF-ADVERTISEMENT)

ReScience

[Re] Interaction between cognitive and motor cortico-basal ganglia loops during decision making: a computational study

Meropi Topalidou[1, 2, 3] **and Nicolas P. Rougier**[1, 2, 3]

1 INRIA Bordeaux Sud-Ouest, Bordeaux, France. **2** LaBRI, Université de Bordeaux, Institut Polytechnique de Bordeaux, Centre National de la Recherche Scientifique, UMR 5800, Talence, France. **3** Institut des Maladies Neurodégénératives, Université de Bordeaux, Centre National de la Recherche Scientifique, UMR 5293, Bordeaux, France.

Nicolas.Rougier@inria.fr

Editor
Tiziano Zito

Reviewers
Benoît Girard
Mehdi Khamassi

Received Jun 9, 2015
Accepted Aug 12, 2015
Published Aug 14, 2015

Competing Interests:
The authors have declared that no competing interests exist.

Code repository

A reference implementation of

→ *Interaction between cognitive and motor cortico-basal ganglia loops during decision making: a computational study*, M. Guthrie, A. Leblois, A. Garenne, and T. Boraud, Journal of Neurophysiology, 109, 2013.

Introduction

We propose a reference implementation of [1] that introduces an action selection mechanism in cortico-basal ganglia loops based on a competition between the positive feedback, direct pathway through the striatum and the negative feedback, hyperdirect pathway through the subthalamic nucleus. The original implementation was made in Delphi (Object Pascal) whose sources are available on request to any of the author of the original article. We have used these sources to disambiguate ambiguous and missing information in the original article. The reference implementation we propose has been coded in Python for ease of reading and Cython for performances because the main result includes a batch of 250 experiments over 120 trials that would be too slow for regular Python scripts.

Methods

We used the description of the model in the original article as well as the sources of the model (requested from author) that are made of a hundred files and 6,000 lines of Delphi for the main source. We have been unable to compile this original implementation but we were able to run the provided Windows executable. We found some factual errors in the original article that have been corrected in this implementation. The initialization of weights are defined in two different parts of the paper. First on page 3030 (second column) *"Weights were initialized to a Gaussian distribution with a mean of 0.5 and a SD of 0.005 at the start of each simulation…"*, then on page 3031 in the caption of figure 4, *"All synaptic weights were initialized to 0.5"*. It happened that both definitions are right but do not address the same projections. Cortico-striatal synaptic weights use Gaussian distribution while all other weights are set to 0.5. Furthermore, the Boltzmann equation given in the original paper uses a . instead of + between first term and second term.

One notable modification in our implementation is the reinforcement learning rule that has been greatly simplified. Original authors have been using quite a complex algorithm for ensuring that *"corticostriatal weights are bounded by a sigmoidal transfer function to represent physical constraints on synaptic growth with an absolute maximum of 0.75 and an absolute minimum of 0.25."*. This algorithm is not described in the article, but from sources, it appears that it is based on the estimation of the weight gradient along the sigmoid. We use instead an Oja-like rule given in the *Synapse* table.

We provide below the formal description of the model according to the proposition of Nordlie et al. [2] for reproducible descriptions of neuronal network models.

Table 1: Model description following [2] prescription.

Table	Description
Populations	Cortex (motor, associative & cognitive), Striatum (motor, associative & cognitive), GPi (motor & cognitive), STN (motor & cognitive), Thalamus (motor & cognitive)
Topology	–
Connectivity	One to one, one to many (divergent), many to one (convergent)
Neuron model	Dynamic rate model
Channel model	–
Synapse model	Linear synapse
Plasticity	Reinforcement learning rule
Input	External current in cortical areas (motor, associative & cognitive)
Recordings	Firing rate & performances

Table 2: Populations

Name	Elements	Size	Threshold	Noise	Initial state	τ
Cortex motor	Linear neuron	1×4	-3	1.0%	0.0	10
Cortex cognitive	Linear neuron	4×1	-3	1.0%	0.0	10
Cortex associative	Linear neuron	4×4	-3	1.0%	0.0	10
Striatum motor	Sigmoidal neuron	1×4	0	0.1%	0.0	10
Striatum cognitive	Sigmoidal neuron	4×1	0	0.1%	0.0	10
Striatum associative	Sigmoidal neuron	4×4	0	0.1%	0.0	10
GPi motor	Linear neuron	1×4	+10	3.0%	0.0	10
GPi cognitive	Linear neuron	4×1	+10	3.0%	0.0	10
STN motor	Linear neuron	1×4	-10	0.1%	0.0	10
STN cognitive	Linear neuron	4×1	-10	0.1%	0.0	10
Thalamus motor	Linear neuron	1×4	-40	0.1%	0.0	10
Thalamus cognitive	Linear neuron	4×1	-40	0.1%	0.0	10
Values (V_i)	Scalar	4	–	–	0.5	-

Table 3: Connectivity

Source	Target	Pattern	Weight	Gain	Plastic
Cortex motor	Thalamus motor	$(1, i) \rightarrow (1, i)$	1.0	0.4	No
Cortex cognitive	Thalamus cognitive	$(i, 1) \rightarrow (i, 1)$	1.0	0.4	No

NEW INCENTIVES

Data citation & Software citation
(recognition, collaboration, track use and re-use, etc.)

Alternative metrics
(Altmetrics, Depsy, PLOS ALMS, Plum analytics, etc.)

New publishers
(F1000, GigaScience, RIO, PeerJ, ReScience, etc.)

European Open Science Agenda
(foster open science, remove barriers, develop infrastructure mainstream access, embed Open Science in society)

Growing concern on reproducibility
(PubPeer, ReScience, p-hacking, etc.)

OPEN HERO

“My work on free software is motivated by an idealistic goal: spreading freedom and cooperation”

GNU founder (1985)
Richard M. Stallman

OPEN HERO

“For me, the repository was supposed to be a three-hour tour, not a life sentence”

ArXiv founder (1991)
Paul Ginsparg

OPEN HERO

“Right now, copyright legislation is so complicated and varies so much from country to country that many people simply choose to ignore it”

European copyright reform (2015)
Julia Reda

OPEN HERO

“Information is power. But like all power, there are those who want to keep it for themselves”

Guerilla Open Access Manifesto (2008)

Aaron Swarzt
1986-2013

OPEN HERO

“My original concept was to provide a free encyclopedia for every single person in the world”

Wikipedia co-founder (2001)
Jimmy Wales

OPEN HERO

“Journal paywalls are an example of something that works in the reverse direction, making communication less open and efficient”

Sci-Hub founder (2011)
Alexandra Elbakyan

SOME VILLAIN

Merck cooked up a phony, but real sounding, peer reviewed journal and published favorably looking data for its products in them. Merck paid **Elsevier** to publish such a tome (2003)

SOME VILLAIN

The publishers **Springer** and **IEEE** are removing more than 120 papers from their subscription services after a French researcher discovered that the works were computer-generated nonsense (2014)

SOME VILLAIN

Wiley is using fake DOIs to trap web crawlers and researchers. **Wiley**'s fake DOIs are not stable, and create the potential for DOI collisions, which break the system (2016)

SOME VILLAIN

Quantitative metrics, grant funding, diversity challenge, gender bias, resistance to change, science careers, etc.

ALL ADA PICTURES FROM
"LOVELAVE & BABBAGE"
BY SYDNEY PADUA

QUESTIONS?
COMMENTS?

OPEN SCIENCE

EDUCATION | ACCESS | SOURCE
REVIEW | METHODOLOGY | DATA